Ontology-Style Web Usage Model for Semantic Web Applications

Thi Thanh Sang Nguyen
Decision Systems & e-Service
Intelligence (DeSI) Lab
Centre for Quantum Computation
& Intelligent Systems (QCIS)
Faculty of Engineering and
Information Technology
University of Technology, Sydney
P.O. Box 123, Broadway, NSW
2007, Australia
tsang@it.uts.edu.au

Hai Yan Lu
Decision Systems & e-Service
Intelligence (DeSI) Lab
Centre for Quantum Computation
& Intelligent Systems (QCIS)
Faculty of Engineering and
Information Technology
University of Technology, Sydney
P.O. Box 123, Broadway, NSW
2007, Australia
helenlu@it.uts.edu.au

Jie Lu
Decision Systems & e-Service
Intelligence (DeSI) Lab
Centre for Quantum Computation
& Intelligent Systems (QCIS)
Faculty of Engineering and
Information Technology
University of Technology, Sydney
P.O. Box 123, Broadway, NSW
2007, Australia
jielu@it.uts.edu.au

Abstract—**Current semantic recommender systems aim to exploit the website ontologies to produce valuable web recommendations. However, Web usage knowledge for recommendation is presented separately and differently from the domain ontology, this leads to the complexity of using inconsistent knowledge resources. This paper aims to solve this problem by proposing a novel ontology-style model of Web usage to represent the non-taxonomic visiting relationship among the visited pages. The output of this model is an ontology-style document which enables the discovered web usage knowledge to be sharable and machine-understandable in semantic Web applications, such as recommender systems. A case study is presented to show how this model is used in conjunction of the web usage mining and web recommendation. Two real-world datasets are used in the case study.**

Keywords-Web Usage Mining (WUM); Web usage model; Web recommendation; Ontology-style model; Ontology-style document

I. INTRODUCTION AND MOTIVATIONS

Semantic Web (SW) technology has emerged recently as the next generation of the World Wide Web that allows future applications to understand and communicate with different knowledge bases [1]. Ontology has been playing an important role in Semantic Web development due to its unique expressive power in modeling a specific domain. Driven not by data structure or format, SW technology builds the meaning (semantics) of the data based on ontology related to the application of interest.

Recommender systems (RS) become more and more important in Web based applications, such as e-commerce, e-government and e-services. These systems generally recommend a set of items which are potentially interesting to a given user [2]. However, current RS require more improvement to make more effective recommendation and to be applicable to various applications [3]. Therefore, RS are advanced into semantic RS which is powered by ontology technology [4]. With the advancement of ontology technology, semantic recommender systems have been proposed [5, 6] to improve web recommendation by exploiting the website ontologies. Ontology is used to represent an application domain for the interference of the meaning of Web pages and the enrichment of Web recommendation with semantic information. However, Web usage knowledge is presented separately and differently from the domain ontology in Web recommender systems. This causes inconsistency between the domain knowledge and Web usage knowledge and results in the complexity of Web recommendation making process. Furthermore, the domain ontology represented by the knowledge of websites does not include the visiting relationships between Web pages which are extracted from usage analysis [7]. These relationships, namely non-taxonomic relationship [8], are less discovered in ontology learning, though it is important and crucial to determine how users visit Web pages. It is desirable to express such Web usage knowledge in a formal representation which is consistent with the domain ontology.

This paper proposes a novel ontology-style model of Web usage. This enables the unification of web usage knowledge and domain knowledge to support semantic web recommendation. This model represents the web access sequences formally in an ontology-style, and output its OWL document for making Web recommendations. The web access sequences can be obtained from a web usage mining (WUM) process [9]. The discovered web usage knowledge is sharable and machine-understandable in semantic recommender systems as well as other semantic Web applications.

The rest of this paper is organized as follows. Section 2 reviews some existing ontology-based recommender systems. Section 3 describes the ontology-style model of Web usage and section 4 presents a case study to illustrate possible applications of this model with some remarks. Section 5 concludes the paper and points out some future works.

II. RELATED WORKS

In recent trend, new generation Web applications are raising using the Semantic Web technology with ontology techniques as the backbone. Ontology can be a good way to model concepts and their relationships in Semantic Web resource development. Hence, most knowledge management systems aim to use ontology as a semantic mean to enhance the sharing of human- and machine-understandable knowledge.

In present ontology-based recommender systems, ontology is often used to enhance the semantics of Web usage data, and is useful for clustering documents, classifying pages or searching subjects. For example, ontology concepts are used to semantically enhance Web logs in a Web personalization system [6], in which one document contains ontology concepts, the document clustering is based on the similarity measure of ontology concepts, and then the system can make recommendations, depending on the input patterns semantically matched with navigational patterns produced in the previous mining phases. Liang Wei and Song Lei [5] construct a website ontology using the domain concepts and significant terms extracted from documents, and generate online recommendations by semantically matching and searching frequent pages discovered from the Web usage mining process. This approach achieves the effectiveness of the precision rate, coverage rate and matching rate.

In case of discovering knowledge from user profiles, ontology can be used to represent the relationships between document categories accessed by users for designing effective Web mining models [10]. Furthermore, ontologies can be discovered from accessed documents to represent the concept models of Web user information needs [11]. In such a case, the built ontology is useful to search right data when a user makes a request. Moreover, the ontology of the discovered knowledge can be expressed in XML and updated for ontology evolution later [11]. However, it is not really efficient to represent user data in the early phases in recommender systems when unexpected data has not been eliminated.

In addition, some works have been undertaken to use ontology based Web mining in hybrid recommender systems with promising results. Ontology can be used to present a set of concepts and their inner-relationships relevant to some knowledge domain [12]. By using this ontology, the website structure and scope can be defined for Web content mining, in which Web pages are concept instances and virtual links between Web pages are relationship instances. This approach shows how well ontology is incorporated into Web mining. On the other hand, according to Trousse, B., et al [7], by using Web usage mining, a domain ontology is able to update the concepts and relations between them, which are derived from usage analysis.

III. ONTOLOGY-STYLE MODEL OF WEB USAGE

The *ontology-style model* of Web usage is an ontology-style representation of the weighted graph of frequent Web access patterns (WAPs) discovered by the web usage mining process [9]. The following subsections give the definition of the model, the representation and implementation of the model, and show how to combine this with the WUM process.

A. Definition

The ontology-style model can be expressed as a 5-tuples: $O=<C, R, P, I, A>$, where

C: Concepts; e.g. Web page is a concept.

R: Relations which are the relationships of visiting Web pages (If a user moves from the current page to the next page, then a relation between the current page and the next page exists. This relation is presented as a non-taxonomic relation [8] "is related to".)

P: Properties; e.g. the attributes of the concept Web page

I: Instances; e.g. *home* page and *history* page

A: Axioms which are a set of rules. For example, one Web page instance is related to zero or many other Web page instances or itself.

The input of this model is the weighted graph of frequent WAPs and its output will be an ontology-style document, which represents the visited web pages and the visiting links among these pages with the corresponding probabilities.

B. Representation of the ontology-style model of Web usage

As defined in subsection A, the ontology-style model of Web usage can be represented in the similar schema as Web ontology expression, namely an ontology-like schema. The concept (*C*) Web page is presented by class Node, including the following properties (*P*):

- *sName*: the name of the Web page,
- *sURI*: the URI of the Web page,
- *fInteresting_rate*: the visiting rate computed based on the number of times that this page was visited.

Properties representing the relationships of visiting the Web pages (*R*) are:

- *sInURI*: the URIs of the previous pages,
- *_outLink*: the next pages.

Class Node describes the Web page as the current page. Users move to the current page from a previous page, and then may go to some pages in the next move with different probabilities. It is worth of noting that the probability of visiting a particular page in the next move from the current page belongs to the combination of the current-page and the next-page, so the property *_outLink* should be represented as an association class OutLink with the following properties:

- *sCurURI*: the linked current page which refers to class Node
- *sOutURI*: the URI of the next page, and
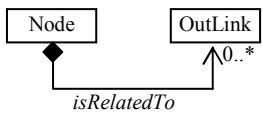- *fOut_prob*: the probability of visiting the next page

C. Implementation

This ontology-style model can be implemented using the commonly used Web ontology language (OWL[1]). Table 1 presents the UML class diagram elements of the ontology-

---

[1] http://www.w3.org/TR/owl2-profiles/

like schema and the corresponding OWL representation. The diagram shows that a Node is composed of zero or more OutLinks. This implies an axiom (𝒜) that is a Web page is related to zero or many other Web pages or itself, so OutLink instances can refer to the next pages, even the same page.

TABLE I.   OWL REPRESENTATION OF WEB USAGE

UML class diagram elements	OWL representation
OutLink	<owl:Class rdf:ID="OutLink"> </owl:Class>
Node — OutLink 0..* isRelatedTo	<owl:Class rdf:ID="Node"> <rdfs:subClassOf> <owl:Restriction> <owl:onProperty rdf:resource="#isRelatedTo"/> <owl:someValuesFrom rdf:resource="#OutLink"/> </owl:Restriction> </rdfs:subClassOf> </owl:Class>
	<owl:ObjectProperty rdf:about="#isRelatedTo"> <rdfs:range rdf:resource="#OutLink"/> <rdfs:domain rdf:resource="#Node"/> </owl:ObjectProperty>

D. Combination with web usage mining process

Figure 1 shows the combination of the ontology-style model of web usage with the web usage mining process [9].

This WUM process is performed by integrating a tree algorithm into the Markov model in order to discover frequent web access patterns from Web log files, and to build a *weighted graph of frequent web access patterns*.
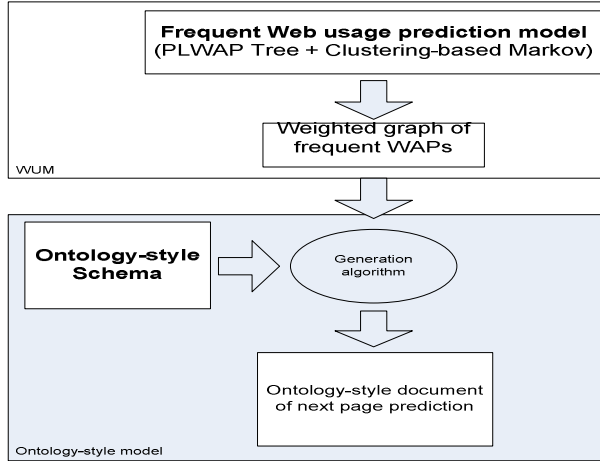

Figure 1. Ontology-style model of web usage with web usage mining process

In this process, the PLWAP-tree algorithm [13] is applied to obtain the frequent sequences in each session of a Web log, and discover all frequent web access patterns (WAPs), in which a frequent sequence is determined based on the measure of support which was used in association rule mining. That means a Web access sequence which has a support that is higher than the predefined minimum support

(MinSup) will become frequent. Moreover, the dynamic clustering-based Markov model [14] is then applied to model the prediction of these frequent WAPs, in which each page is represented by a state. With such a model, the prediction of the next page depends on the current page and the previous pages. In some cases, a state may have a few clones as other prediction cases. Each clone describes a case of next page prediction depending on which previous pages are clustered to guarantee a second-order probability accuracy. As a result, a weighted graph of frequent WAPs is built from the model.

The above approach achieves the benefits of the PLWAP-tree and the dynamic clustering-based Markov model and overcomes their drawbacks by omitting uninteresting web pages. According to C.I. Ezeife and Y. Lu [13], the PLWAP-tree algorithm was verified by comparing with Apriori-like algorithms, so frequent WAPs which will be discovered in this paper are reliable. The dynamic clustering-based Markov model which is an advanced probabilistic model of the Markov model contributes to computing the second-order probability of visiting Web pages in frequent WAPs. However, the results generated by this process are only human-understandable and hard to be shared in Semantic Web applications because the representation of the resultant graph is lack of formal semantic expressive power.

Given the OWL classes and the weighted graph of frequent WAPs, an ontology-style document of web usage will be automatically generated to represent the web usage knowledge in a formal way using OWL.

IV.   A CASE STUDY

This case study shows the possible application of the proposed ontology-style model of web usage in a simple web page recommendation system.

A. Flowchart of a web recommender system

Figure 2 shows the flowchart of the web recommendation system.

As shown in Figure 2, firstly, the input data of the model is discovered from the previous WUM process [9]. Secondly, the ontology-style model is built to generate a document of Web usage. An ontology building tool, Protégé (http://protege.stanford.edu/doc/dev.html), is used to implement the OWL class instances. Thirdly, the built document will be used for making web recommendation.
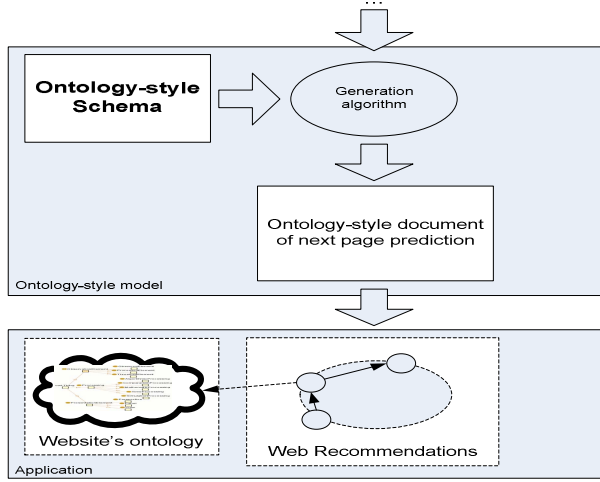
Figure 2. Flowchart of Web recommendation using the ontology-style model

B. User session datasets

In this case study, we considered two real-world live Web sites: NASA Web site (http://science.ksc.nasa.gov), and ClarkNet Web site (http://ita.ee.lbl.gov/html/contrib/ClarkNet-HTTP.html). The NASA access session dataset was collected from 00:00:00 August 1, 1995 through 12:16:12 August 10, 1995, a total of 10 days. The ClarkNet access session dataset was collected from 00:00:00 September 4, 1995 through 23:59:59 September 10, 1995, a total of 7 days. These datasets are obtained after cleansing the web logs, i.e., removing erroneous and invalid pages from the web logs and eliminating multi-media files such as gif, jpg and script files such as js and cgi. Table 2 shows the size of the datasets after preprocessing.

TABLE II. DATASETS AFTER PREPROCESSING

Dataset	Number of users	Number of Web pages	Number of sessions
NASA	26,037	1,446	49,406
ClarkNet	82,318	10,042	122,674

C. Ontology-style documents

Each dataset of user sessions was mined through the frequent Web usage mining module [9] to obtain the weighted graph of frequent WAPs, which will then be fed into the proposed ontology-style model of web usage to generate the ontology-style document. For example, Figure 3 shows part of the generated ontology-style document of ClarkNet website usage. Tag *sURI* and *sName* describe the URI and name of the Web page *vendela2*. Tag *_outLink* specifies *OutLink* individuals including the URI of a next page and the corresponding visiting probability. Tag *sInURI* describes the URI of a previous page of the page *vendela2*.

```
<Node rdf:about="#133_1.vendela2">
    <_outLink>
        <OutLink rdf:about="#133_1.vendela2.64.swim">
            <sOutURI
                rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
                http://www.clark.net/pub/job/swim/swim.html
```

```
            </sOutURI>
            <fOut_prob
                rdf:datatype="http://www.w3.org/2001/XMLSchema#float">
                0.5
            </fOut_prob>
        </OutLink>
    </_outLink>
    <_outLink>
        <OutLink rdf:about="#133_1.vendela2.End">
            <sOutURI
                rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
                End
            </sOutURI>
            <fOut_prob
                rdf:datatype="http://www.w3.org/2001/XMLSchema#float">
                0.5
            </fOut_prob>
        </OutLink>
    </_outLink>
    <sURI
        rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        http://www.clark.net/pub/job/vk/vendela2.html
    </sURI>
    <sInURI
        rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        http://www.clark.net/pub/job/vk/vendela.html
    </sInURI>
    <sInURI
        rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        Start
    </sInURI>
    <sName
        rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        vendela2
    </sName>
    <fInteresting_rate
        rdf:datatype="http://www.w3.org/2001/XMLSchema#float">
        4.0
    </fInteresting_rate>
</Node>
```

Figure 3. Node describes page *vendela2.html* in case of ClarkNet dataset (*MinSup=0.001*)

Table 3 tabulates the number of Nodes (Web pages) and the number of OutLinks in the ontology-style documents discovered from the NASA and ClarkNet dataset for different minimal support (MinSup) values. It can be seen from Table 3 that there are more Web pages discovered with lower *MinSup* value. In order to achieve a better prediction result, we can set a lower *MinSup* value to increase the number of frequent Web pages discovered.

TABLE III. NUMBER OF NODES AND OUTLINKS

Dataset	MinSup	Number of generated Nodes	Number of generated OutLinks
NASA	0.01	50	66
	0.005	88	183
	0.001	845	5590
ClarkNet	0.01	18	21
	0.005	52	67
	0.001	361	5435

D. Web recommendation

Given the ontology-style document, the next-page recommendation can be generated based on the following reasoning algorithm for the likely-visited Web pages from the current page

Input: current-page = A and previous-page = B
Loop Each Node in the ontology-style document *do*
If a Node describes web-page A *and* previous-page B, *then*
return
- All next pages of page A
  recommend

> - The next page with the highest visiting probability
> *end loop*
>
> where, A and B can be page names or URIs

E. Implementation

This reasoning algorithm is implemented using the Protégé-OWL API. By applying the ontology-style reasoning to the ontology-style document, the machine can understand which pages are related to a given page and be able to generate the likely-visited next page with probability. Two examples are illustrated as following.

Example 1: NASA dataset with *MinSup=0.01*, recommend the next page for the current page *history*

When the page *missions* or *ksc* was previously visited, the recommendation is to go to the page *apollo*. The reasoning process is as shown below.

> *If* previous page:
> - http://science.ksc.nasa.gov/shuttle/missions/missions.html OR
> - http://science.ksc.nasa.gov//ksc.html
> *And* current page: http://science.ksc.nasa.gov/history/history.html
> *Then* the following pages may be visited with the corresponding probabilities.
> http://science.ksc.nasa.gov/history/apollo/apollo.html (40%)
> http://science.ksc.nasa.gov/shuttle/missions/missions.html (20%)
> End: No visit more. (40%)
> → Recommendation: http://science.ksc.nasa.gov/history/apollo/apollo.html

Example 2: ClarkNet dataset with *MinSup=0.005*, recommend the next page for the current page *vendela2*

When the page *vendela* was previously visited, the recommendation is to go to the page *swim*. The reasoning process is shown below.

> *If* previous page:
> - http://www.clark.net/pub/job/vk/vendela.html
> *And* current page: http://www.clark.net/pub/job/vk/vendela2.html
> *Then* the following pages may be visited with the corresponding probabilities.
> - http://www.clark.net/pub/job/swim/swim.html (50%)
> - End: No visit more (50%)
> → Recommendation: http://www.clark.net/pub/job/swim/swim.html

The recommendation results from examples 1 and 2 have been manually verified against the weighted graph of frequent WAPs discovered by the web mining process [9].

F. Possible applications

Since the ontology-style model and document are implemented using OWL, we can easily incorporate this ontology-style model with a ontology-based recommender system which used ontology to describe the meaning of Web pages [5], so that the knowledge from Web usage mining and website ontology can be unified. In such a way, the Web recommendation system can understand not only what the page is about, but also where the page is in Web usage relationships and this can make better recommendations. This would significantly simplify the existing recommendation making process [5, 6], which make many efforts to combine ontology and Web usage mining.

In addition, this ontology-style model can be used in enriching *domain knowledge representation*. Given the ontology-style document of the Web usage knowledge, a program can automatically determine a Web page and related pages. From another hand, a Web page can be described by concepts in a domain ontology [5]. By analyzing the relation attributes "*is related to*" of the ontology-style model, the program can find the relations between the concepts of related pages and automatically add these relations into the domain ontology. For example, we assumed that the domain ontology of website NASA and ClarkNet were built. Based on the previous results:

- Example 1 shows that pages *history* and *missions* are related, so the relation between *history* and *missions* need to be considered and added into the domain ontology of website NASA.
- Example 2 shows that pages *vendela* and *swim* are related, so the relation between *vendela* and *swim* need to be considered and added into the domain ontology of website ClarkNet.

This enrichment of the domain ontology requires some extensions of the proposed ontology-style model of web usage, such as adding some attributes of the concept Web page to create connections to the domain ontology, and adding relations between concepts into the domain ontology.

Furthermore, the ontology-style documents of Web usage can become external Web resources which allow other applications reusing and extending more attributes to enrich the domain ontology with Web visiting relationships.

V. CONCLUSIONS AND FUTURE WORKS

In this paper, we propose an ontology-style model of Web usage and show its promising applications. The model represents the weighted graph discovered from raw web logs using the ontology language OWL. The model can capture the non-taxonomic relations between the visited pages discovered from web usage mining for supporting semantic Web recommendation and enrichment of domain ontology.

Compared with the weighted graph, this proposed ontology-style model achieves substantial benefits which are machine-understandable, sharable on Web, and capable of ontology reasoning. As experimented, it is easy to interpret ontology-style documents via the defined ontology schema for recommendations in ontology-based systems.

The proposed ontology-style model describes successfully the relationships of visiting Web pages which are the core knowledge of Web recommendation in semantic recommender systems.

In the future, the proposed approach might be adopted to develop some Web 3.0 applications to support users' decision making in Web browsing process.

REFERENCES

[1]   M. Hepp, "Ontologies: State of the Art, Business Potential, and Grand Challenges," *Ontology Management: Semantic Web, Semantic Web Services, and Business Applications*, Springer, Heidelberg, 2008, pp. 3-22.

[2]   B. Mobasher, *Recommender Systems*, 2007.

[3]   G. Adomavicius and A. Tuzhilin, "Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Trans. on Knowl. and Data Eng.*, vol. 17, no. 6, 2005, pp. 734-749; DOI http://dx.doi.org/10.1109/TKDE.2005.99.

[4]   M. Adda, et al., "Toward Recommendation Based on Ontology-Powered Web-Usage Mining," *Internet Computing, IEEE*, vol. 11, no. 4, 2007, pp. 45-52.

[5]   L. Wei and S. Lei, "Integrated Recommender Systems Based on Ontology and Usage Mining," *Active Media Technology*5820, Springer-Verlag Berlin Heidelberg, 2009, pp. 114–125.

[6]   M. Eirinaki, et al., "Introducing Semantics in Web Personalization: The Role of Ontologies," *Semantics, Web and Mining*, Springer-Verlag Berlin Heidelberg, 2006, pp. 147 – 162.

[7]   B. Trousse, et al., "Web Usage Mining for Ontology Management," *Data Mining with Ontologies: Implementations, Findings, and Frameworks, 2008*, H. N. S. C. D. Xodo, ed., Information Science Reference, 2008, pp. 37-64.

[8]   D. Sánchez and A. Moreno, "Discovering Non-taxonomic Relations from the Web," *Proc. 7th International Conference on Intelligent Data Engineering and Automated Learning*, 2006, pp. 629-636.

[9]   S.T.T. Nguyen, "Efficient Web Usage Mining Process for Sequential Patterns," *iiWAS2009*, ACM, 2009 of Conference, pp. 455-459.

[10] Y. Li and N. Zhong, "Ontology-based Web mining model: representations of user profiles," *Proc. Proceedings of the IEEE/WIC International Conference on Web Intelligence (WI 2003)*, 2003, pp. 96-103.

[11] Y. Li and N. Zhong, "Mining Ontology for Automatically Acquiring Web User Information Needs," *IEEE Transactions on Knowledge & Data Engineering*, vol. 18, no. 4, 2006, pp. 554-568.

[12] R.M. Suresh, "A study on the ontology based Web mining for digital library," *Proc. Information and Communication Technology in Electrical Sciences (ICTES 2007), 2007. ICTES. IET-UK International Conference on*, 2007, pp. 1096-1100.

[13] C.I. Ezeife and Y. Lu, "Mining Web Log Sequential Patterns with Position Coded Pre-Order Linked WAP-Tree," *Data Mining and Knowledge Discovery*, vol. 10, no. 1, 2005, pp. 5-38; DOI 10.1007/s10618-005-0248-3.

[14] J. Borges and M. Levene, *A Dynamic Clustering-Based Markov Model for Web Usage Mining*, Technical Report, Available online at http://xxx.arxiv.org/abs/cs.IR/0406032, 2004.